# Project Report:

**Overall Summary:**

One of the open data sources that the CSV file was obtained from was NYC open data that is related to the NYPD's arrest data with the description of what the individual is being arrested for, which five boroughs it happened at, date of arrest, etc. In addition to the CSV information, it is last updated up until September 2023. In my findings/research, I would like to know among the five boroughs I wonder which seems to have the most crimes being committed. In this project, I will be creating heatmaps - one with the imports of folium and one without the imports of folium. In the creation of the heat map, I will create a pivot table by combining two data sets (age group, borough, and unfiltered crime/offense description). Furthermore, I will import folium in the creation of my heatmap for a visual view (OpenStreetMap) comparing unfiltered versus filtered crime data. I will then create a bar plot with the filtered crime activity to my preference to determine what age groups are committing what is what. Lastly, we will then have a time series plot to see the overall trend of the crime categories of the crime count that is committed monthly.

**Datasets to be used from the CSV files:**

1) ARREST_DATE
2) OFNS_DESC
3) AGE_GROUP
4) ARREST_BORO

**CSV Files Used via NYC Open Data:**

'NYPD_Arrest_Data_Year_to_Date_.csv'

https://catalog.data.gov/dataset/nypd-arrest-data-year-to-date

**Data may vary depending on the source update**

**Note:** Data are most recent for the year of 2023 - Goes up until the month of September of 2023

**Introduction/Approach:**

The CSV file that I will be using consists of datasets that will include criminal charges description, arrest in which borough of the individual, etc. Descriptions of each criminal activity/action may include low to high in terms of danger and the violation of the law. Furthermore, limiting these activities will maintain a safer environment in overall settings. The latest CSV file I will be using (up until September of 2023) is from NYC open data. I will examine the offense by creating data visualizations. In my findings, I will be trying to determine how these categorized offenses played a role within the five boroughs and will be figuring out which borough has the highest crime rate.

**Background (Problem Statement):**

Crimes and or other categorized offenses can be ranked from low to high in terms of danger. Wouldn't you want to live in a peaceful and safe environment? Where would you want to reside based on your knowledge about the crime rates and other criminal activity that has happened (its history) within the five boroughs? I will be testing and filtering out my preference for certain criminal/offenses for each borough and go from there.

**What Will I Be Experimenting/Testing With (please see google colab - link provided):**

In the findings and testing for the experiment, I will create two heat maps (one with folium & one without). For the two heat maps, I will not be filtering out any criminal activity description provided in the CSV and will move on from there. I will then create another heatmap - using folium but this time I will be filtering the criminal activity/offenses to my desire and compare my findings to see whether there are any differences and or similarities.

In addition to this, I will be also creating a bar plot with filtered criminal activities and have it combined with age groups to see what group of ages are often committing these activities.

Finally, I have created a time series plot. In the making of this I am interested In knowing does each month the crime rate will lower or will it increase.

**Approach (please see google colab - link provided):**

Creating a heatmap will allow me to visually see based on the color which crimes were done the most. In addition to this, I have combined two data sets to my findings (OFNS_DECS & AGE_GROUP) - which correspond to criminal activity record/description and the age of the person that has committed the activity.

In addition to this, I have created another heat map with the imports of folium to visually view it on an open street map. This allows the audience to visually see the boroughs that have the highest crime committed.

**Note:** The longitude and latitude are based on the entire borough and not a specific part

After the unfiltered data, I decided to filter certain criminal activity/offenses to my preference to see if there were any differences in comparison to filtered and unfiltered data. (After doing a comparison there seem to be no differences and results seem to be the same as unfiltered)

To deepen my findings in knowing what age groups committed these types of activities with the filter of the offenses to my preferences, I created a bar plot. The bar plot that I have shown which age group has committed the most crimes and what crimes are done the most.

Lastly, a time series plot - here it allows you to see overtime of each criminal activity done through the months of 2023 (most recent) in addition to this the time series plot gives you a visual view of whether these crime activity trend is increasing or decreasing over time.

**Results/Findings (please see google colab - link provided):**

In my findings ~ with unfiltered data sets (ARREST_DATE, OFNS_DECS, AGE_GROUP, ARREST_BORO) I found out that within the spread-out age group - ages from 25-44 commit the most criminal activity in the borough of Brooklyn (Kings).

In the creation of a heat map with the imports of folium (filtered vs unfiltered), the results are the same. Whereas the most committed crimes are mostly located in Brooklyn (kings). In addition, the conclusion that can be drawn by looking at the OpenStreetMap we can see that in the borough of Staten Island, it seems that there are fewer crime activities.

Creating a bar plot combining both the age groups and criminal activity of preference, I wanted to see what the most committed crimes are among the ages. My findings were ages 25 through 44 played a huge chunk of assaults, robberies, and dangerous weapons in position.

With the time series plot, we can visually see the trend of the crimes that I chose that are committed through each month from the start of January to September of 2023. In conclusion, by looking at each individual plot Brooklyn (kings) seems to be on top and has a slight increase towards the end of September in overall view. In addition, by looking at the time series plot we can say that the borough of Staten Island has the lowest rate of crime when compared to the other four boroughs, but does it have to do with the population?

**Conclusion:**

In conclusion to my findings for the year 2023 up until September, filtered or unfiltered data sets we can say that Kings - Brooklyn has the most crimes where it is committed. Although we aren't focused on one area or place, we still can say that Brooklyn (kings) is where most of the criminal activity has been committed/done.

Having these data all stored allows data scientists to recreate plots - to visually understand what is happening based on combining multiple data sets to come up with a result/conclusion to draw from and how are crimes like in each borough.

In addition to this, in my overall findings it seems that the borough Staten Island seems to have the most minimum and where Brooklyn has the maximum crimes has committed.

Note: Based on what I see

High to Low in term of crime committed (filtered data sets - use of time series plot) -

1 - Brooklyn (kings)
2 - Bronx
3 - Manhattan
4 - Queens
5 - Staten Island